Modelling cumulative exposure for inference about drug effects

Modelling cumulative exposure for inference about drug effects in observational studies

Running title: Modelling cumulative exposure for inference about drug effects

Bassam Farran[1], Stuart McGurnaghan[1], Helen C Looker[2], Shona Livingstone[2], Eva Lahnsteiner[2], Helen M Colhoun[1, 3], Paul M McKeigue[4]

[1]Institute of Genetics and Molecular Medicine, University of Edinburgh; [2]University of Dundee; [3]NHS Fife; [4]Usher Institute for Population Health Sciences and Informatics, University of Edinburgh

Corresponding Author:
Helen Colhoun
AXA Chair in Medical Informatics and Life Course Epidemiology
Institute of Genetics and Molecular Medicine
Western General Hospital
Crewe Road, Edinburgh, EH4 2XU, UK
helen.colhoun@igmm.ed.ac.uk

Keywords: cardiovascular disease, CVD, modelling cumulative exposure, type 2 diabetes, statins exposure, observation study

Key Points:

- Observational studies are necessary to evaluate rare or long-term drug effects that cannot be assessed in clinical trials
- Models need to be specified correctly to avoid confounding and allocation bias
- Including ever-vs-never exposure to drugs together with cumulative exposure helps control these issues
- We demonstrate this using statin exposure on Cardiovascular disease, as statin effects on CVD are well known
- We also provide a mathematical explanation of how modelling cumulative and ever-exposure jointly controls for time-invariant allocation bias

Main text word count: 2985

Conflict of Interest Statement: There are no prior postings and presentations of this paper. The authors declare no conflict of interest. The Scottish Diabetes Research Network receives financial support from NHS Research Scotland (NRS). Financial support for programmers' time was provided by the Chief Scientist Office (CSO) of the Scottish Government.

Abstract

Purpose: To demonstrate a modelling approach that controls for time-invariant allocation bias in estimation of associations of outcome with drug exposure.

Methods: We show that in a model that includes terms for both ever- versus never-exposure and cumulative exposure, the parameter for ever-exposure represents the effect of time-invariant allocation bias and the parameter for cumulative exposure represents the effect of the drug after adjustment for this unmeasured confounding. This assumes no stepwise effect of the drug on the event rate, no reverse causation, and no unmeasured time-varying confounders. We demonstrated this by modelling the effect of statins on cardiovascular disease, for which the true effect has been well-characterized in randomized trials, using time-updated Cox regression models in a national cohort of Type 2 diabetes patients.

Results: The crude hazard ratio associated with ever-use of statins was 1.13 in a standard cohort analysis comparing exposed with unexposed person-time intervals. When ever-never use and cumulative exposure are modelled jointly, the effect of statins can be estimated from the cumulative exposure parameter (hazard ratio 0.97 per year of exposure, 95% CI 0.97 to 0.98). The ever-exposed term (hazard ratio 1.20, 1.16 to 1.23) in this model can be interpreted as estimating the allocation bias.

Conclusions: Where stepwise effects on the risk of adverse events are unlikely, as for instance for effects on risk of cancer, joint modelling of ever-never and cumulative exposure can be used to study the effects of multiple drugs and to distinguish causal effects from confounding by allocation.

Introduction

Clinical trials are usually not of sufficient size or duration to capture all clinically important adverse or unexpected beneficial effects of drugs. Thus observational pharmaco-epidemiological studies remain important as a means of evaluating drug safety.[1,2] Such observational studies have a range of design and interpretation challenges one of the most important being termed confounding by indication or 'allocation bias'. Such confounding can occur when simple comparisons of rates of outcomes are made between those exposed (hereafter ever-users) and those not exposed to a drug (hereafter never-users) at some point in time. In these comparisons, ever-users and never-users of the drug may differ in susceptibility to the outcome in question for reasons independent of the drug. In other words, *propensity* to be prescribed the drug is associated positively or negatively with the outcome in question. Such confounding by indication or allocation bias cannot reliably be dealt with by simple adjustment for measured covariates such as past medical history because there are subtle unmeasured confounders such as co-morbidity that are not fully captured by the information in medical records.[3] Propensity score matching, which matches exposed and unexposed individuals on covariates that predict drug allocation, does not overcome this limitation of relying on measured covariates.[4] One approach to dealing with such fixed between-person confounding by indication or allocation bias that can be useful for assessing certain drugs and outcomes of interest is to focus inference about drug effects on whether the outcome changes with *cumulative exposure* - a well-known concept in epidemiology that is enshrined in the dose-response criterion of Bradford Hill's criteria for causality [5] - and to contrast person-time-intervals with high cumulative exposure with person-time intervals with low (but nonzero) cumulative exposure.

Investigators have used various approaches to modelling cumulative exposure effects of drugs. Typically models use categorical terms for cumulative exposure with the reference category being intervals or persons in which no exposure has occurred. We and others [6] have used models where

time updated terms for ever versus never exposure and cumulative exposure are included simultaneously in the model. This has the advantage of permitting evaluation of cumulative exposure whilst allowing simultaneous evaluation of cumulative effects of other drugs and use of all covariate data. We have used this approach to show for example that thiazolidinediones are associated with increased risk of hip fracture [7], and that there was no detectable effect of pioglitazone with bladder cancer.[8]

In this paper we explore further the characteristics of such cumulative effects models, what allocation related biases they help overcome and which sorts of allocation-related bias they do not avoid. To do this, we use data on statin exposure and cardiovascular disease to illustrate the different estimates that are obtained when simple models of ever versus never exposure are used compared with models of cumulative exposure since the effects on cardiovascular disease (CVD) have been demonstrated unequivocally in randomized trials [9–12] and since it is expected that there will be a difference in the prior risk of CVD in those prescribed versus not prescribed statins. We emphasise of course that studies of association of drug exposure with outcomes are not sufficient to prove causality but are useful along with other considerations of causality in assessing effects of drugs in observational studies. Since they are commonly used,[13–19] our exploration of such modelling is of practical relevance to interpreting observational pharmaco-epidemiological studies.

Methods

Data source

We used data from a previously described diabetes register in Scotland (SCI-Diabetes) that captures all issued prescriptions (but not prescription encashments) for any diagnosis and other covariate data on the total population with diabetes since 2004.[20] These data were linked using the national

unique health care identifier, the Community Health Index number, to the national hospital admissions dataset and to the national death register which extend back to 1981.

Inclusion criteria, entry and exit times

All those aged at least 18 years when diagnosed with type 2 diabetes and in the register at any point between 01/01/2004 (study start date) and 31/12/2011 (study end date) were included. Those with a CVD event prior to entry date, as identified from the hospital admissions database (see definition below), were excluded. The entry date into this study depends on diabetes diagnosis date: if the patient was diagnosed prior to study start date, they entered the study when they first became observable for drugs (an individual was considered to be observable for drugs when they received at least one prescription of any kind in a six month period) on or after study start date. If they were diagnosed after study start date, they entered the study when they first become observable for drugs on, or after their diabetes diagnosis date. All such persons are fully observable for national hospital admissions and deaths. The exit date for each person was the earliest of: first CVD event, death, study end date, or ceasing to be evaluable for prescribing and admission data i.e. exit from the country. Individuals who were already taking a statin at the time they first came under observation for drugs were excluded because their cumulative exposure was unknown. Any individual contributing less than 28 days to the study was excluded. In total we included 982,571 person-years at risk with 626,667 accruing for ever-exposed and 355,904 for never-exposed.

Events

CVD events were defined as hospital admission or death with any of the following ICD9 codes (410-414, 431, or 433-435) or ICD10 codes (I20–I25, I61, I63, I64, or G45). During the study there were 32,201 events, 8,573 (26.6%) among never-users and 23,628 (73.4%) in ever-users.

Data preparation for modelling and calculation of exposure

The period of follow-up was divided into discrete intervals of equal length to allow time-updating of covariates.[21] For initial analyses, a 56-day interval length was used. Shortening this to 28-day

intervals changed the effect estimates only slightly in the third significant digit: this 28-day interval was therefore taken to be sufficiently close to the theoretical limit of a model with infinitesimally short time intervals. A data matrix was generated with one row for each person-time interval during which the individual was under observation, and columns specifying event status at end and drug exposure and covariate values at the beginning of each 28 day interval. Cumulative exposure in days at the start of each interval was calculated as a sum over all earlier intervals in which a statin prescription was current. The exposure duration was calculated as the sum of drug eras, where each drug era is a sequence of prescriptions where the issue date does not exceed 56 days from the end of the previous prescription. This cut-off for defining drug eras has been validated (for other drugs) by using the directions for use (dosage and frequency) and the total quantity dispensed to calculate the durations of prescriptions. Gaps between drug eras were not included in exposure time. Following drug cessation the cumulative exposure was carried forward in subsequent time intervals. Covariates were sex, time-updated age at beginning of interval, age at diabetes diagnosis, calendar time, smoking (ever/never), low density lipoprotein (LDL) cholesterol, systolic blood pressure, diastolic blood pressure, glycosylated haemoglobin A_{1c} (HbA_{1c}), estimated glomerular filtration rate (eGFR), body mass index, ever-exposure to statins (defined to take values 1, 0 according to whether the person has ever been exposed to statin before or during that person-time interval) and cumulative exposure to statins (defined as cumulative years of statin use). The ever-exposed and cumulative exposure variables are not collinear as long as at least one individual is exposed for at least two 28-day intervals.

Statistical methods

Cox regression models for time to failure were specified to include statin exposure as (i) only the time-updated ever-never exposure term (ii) both the ever-never term and a linear term for cumulative exposure and (iii) cumulative exposure coded as a categorical variable, with baseline category defined as either (a) unexposed person-time intervals or (b) exposure greater than zero but

less than 0.5 years. A mathematical explanation of how modelling cumulative exposure and ever-exposure jointly controls for time-invariant allocation bias is given in the online supplementary material.

Results

To show how the association of statin exposure with CVD is confounded, Table 1 compares CVD risk factor status in ever-users compared to never-users as of study mid-point (July 1st 2007). Levels of most CVD risk factors were higher in ever-users than in never-users. Table 2 shows the results of fitting Cox proportional hazards models in which statin exposure is coded as time-updated ever-use, comparing models with and without an additional term for cumulative exposure. With statin exposure coded only as ever-use, as in a standard cohort analysis of a time-varying exposure, statin exposure is associated with a higher event rate. Adjusting for additional covariates - smoking status, body mass index, total cholesterol, HbA_{1c}, eGFR, systolic and diastolic blood pressure, and use of anti-hypertensive drugs - reduces the size of the hazard ratio parameter from 1.13 to 1.05, but does not reverse its direction. If naively interpreted, this would suggest an effect of statin use on CVD opposite in direction to that demonstrated by many clinical trials.[22] This illustrates that adjusting for the risk factors that are typically available in an electronic health record is not enough to control for confounding by indication or allocation bias.

Adding a linear term for cumulative exposure to statins to this model increases the hazard ratio associated with ever-use of statins from 1.13 to 1.20, but shows an inverse relationship of cardiovascular risk to cumulative exposure that is consistent in direction with the protective effect that has been demonstrated in trials. Adjustment for additional covariates changed the hazard ratio for cumulative exposure only slightly from 0.97 to 0.96 per year of exposure.

The form of the relationship of cardiovascular risk to cumulative exposure to statins is examined further in Table 3, which summarizes two alternative models in which cumulative exposure is coded as a categorical variable. In columns two to five, the model encodes cumulative exposure with unexposed person-time intervals as the baseline category, corresponding to a standard analysis of a cohort study to evaluate cumulative exposure.[23,24] With this model, those exposed to statins appear to have increased risk for CVD initially, which is inconsistent with the trial evidence that the risk of CVD begins falling from the start of statin usage. In columns six to nine of Table 3, unexposed person-time intervals are excluded and the baseline category for cumulative exposure is the first *exposed* category (up to 6 months). With this model, the direction of the cumulative effect is consistent with trial evidence. The same hazard ratios can be generated from a dataset that includes unexposed person time-intervals if this model includes a time-updated term for ever-exposure and up to 6 months (including unexposed person-time intervals) as the baseline category for cumulative exposure.

To make this clear without resorting to statistical modelling, Figure 1 shows the age-standardised rates for CVD by cumulative statin exposure. The highest rates are for those in the initial period of exposure to statins, followed by a decline. This is consistent with allocation of higher-risk individuals to statin, and a protective effect of the drug.

Discussion

In pharmacoepidemiology, confounders of the association between drug exposure and outcome include subtle factors influencing drug allocation that are not recorded in electronic health records. We have demonstrated that evaluating statin effects simply by comparing rates in exposed and

unexposed person-time intervals would lead to erroneous conclusions about statin effects on CVD, as the rate of cardiovascular events is higher in exposed than in unexposed person-time intervals. This is to be expected, as current guidelines for statin prescribing emphasize that they are indicated in patients at high risk of cardiovascular disease. Even where additional efforts are made to collect data on and adjust for confounders, for instance in studies of the benefits and risks associated with post-menopausal oestrogen replacement therapy,[25] it is still difficult to eliminate the allocation bias that results from unmeasured confounding. By contrasting person-time intervals with high and low cumulative exposure among those who have been prescribed the drug at least once, we can eliminate time-invariant confounding by factors associated with allocation (i.e. fixed allocation bias).

One might ask why one cannot simply evaluate drug effects by comparing the rate of the event of interest in users prior to the start (of drug use) with their rate as their use accumulates. The reason is that for many exposures and outcomes of interest it is necessary to restrict the study to incident first events because the event itself may alter the probability of subsequent exposure. One might also ask why one cannot just measure the rate directly at the time of initiation of the drug and then assess whether it subsequently changes. The reason is that unfeasibly large sample sizes would be needed to give a reliable estimate of the ratio of the rate in this short time interval at the start of exposure to the unexposed rate. Instead the rate at the time of initiation of exposure is estimated by extrapolating back to zero the line (or curve) fitted for the relation of risk to years of cumulative exposure. The ratio of this estimated rate to the rate in unexposed person time intervals corresponds to the parameter for ever-exposure. If the time course of the effect of the drug on the hazard rate has been modelled correctly (as a line or curve), this parameter for ever-exposure represents the allocation bias. This is an example of the more general principle that the crude rate ratio can be

factored into a parameter that measures the strength of confounding and an adjusted parameter that measures the effect of exposure.[26]

A simpler approach is to limit all analyses to exposed person-time intervals, and to estimate cumulative effects among these exposed individuals. However, such an approach does not fully use all the information on effects of other covariates on the outcome that may be available in the full dataset. These effects include calendar time effects which may be closely correlated with cumulative exposure. Most importantly, the exposed-only approach can be used only to study one drug at a time, and cannot be used to model the joint effects of several drugs.

For the regression coefficients for ever-never and cumulative exposure to be interpreted as parameters that measure respectively the strength of confounding and the causal effect of the exposure, several assumptions are required. First, it is assumed that there is no stepwise effect of the drug on the event rate, so that the event rate in individuals who have just started the drug can be estimated by extrapolating the relationship (not necessarily linear) of hazard rate to cumulative exposure back to zero. This is a reasonable assumption for example when the adverse event under study is cancer (since tumours take many years to grow), or osteoporotic fractures (though an increase in fracture risk could result from a stepwise effect on falls rather than a cumulative effect on bone density. For cardiovascular disease, the plausibility of assuming no stepwise effect depends upon whether the drug effects are on the underlying atherosclerosis, which develops and regresses only slowly, or on more immediate triggers such as plaque rupture and thrombosis. For evaluating drug effects such as acute liver toxicity that might be expected to occur quickly after exposure, a cumulative exposure term would be hard to interpret. Second, it is assumed that the allocation bias is time-invariant. For instance, if higher-risk patients are more likely than lower risk patients to

discontinue the drug in the first few months because of adverse effects, estimation of the cumulative exposure term would falsely indicate a protective effect. Third it is assumed that there is no reverse causation, as for instance where a disease process causes a drug to be prescribed before the underlying disease is diagnosed. In this situation, incidence of the disease will be high shortly after the time of initiation of the drug and will fall with increasing cumulative exposure. An example of this is the association between exposure to insulin and pancreatic cancer.[27]

Large scale observational pharmaco-epidemiological studies are becoming increasingly feasible as large electronic databases of drug exposures and health outcomes become available. Several scientists have written elegantly about the perils of such studies and the ease which errors in design can be made (26–30). For example, Suissa has emphasised that incorrect design and analytic methods in such studies frequently gives rise to immortal-time bias, which can be viewed as a differential misclassification of exposure status in relation to time to event.[28–30] Unfortunately the analyses in many published studies in this field are still affected by some form of confounding or bias, limiting the conclusions that can be drawn despite the wealth of data available (27–29). Van Staa et al demonstrated the issues relating to allocation bias in their study of diabetes treatment and cancer risk.[31] Thus, improving the methodology for analysing these data is vital. In this paper we aimed to provide some useful insights to those considering modelling cumulative effects as one of the possible means to reduce the potential for allocation bias. We emphasise that associations between drug exposures and outcomes are only one element to consider in assessing the evidence for causality. Consideration must be given to other factors including consistency of effects in different settings and biological plausibility.

Acknowledgements

Additional Contributions: We thank the members of the Scottish Diabetes Research Network and the SCI-DC team.

Financial Support: The Scottish Diabetes Research Network receives financial support from NHS Research Scotland (NRS). Financial support for programmers' time was provided by the Chief Scientist Office (CSO) of the Scottish Government.

Role of the Funder/Sponsor: The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

References

1. Hennessy S. Use of health care databases in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol* 2006; **98**: 311–313. doi:10.1111/j.1742-7843.2006.pto_368.x.

2. Hashikata H, Harada KH, Kagimura T, Nakamura M, Koizumi A. Usefulness of a large automated health records database in pharmacoepidemiology. *Environ Health Prev Med* 2011; **16**: 313–319. doi:10.1007/s12199-010-0201-y.

3. Yang X, Kong AP, Luk AO, *et al.* Validation of methods to control for immortal time bias in a pharmacoepidemiologic analysis of renin-angiotensin system inhibitors in type 2 diabetes. *J Epidemiol Jpn Epidemiol Assoc* 2014; **24**: 267–273.

4. Wang J, Donnan PT. Propensity score methods in drug safety studies: practice, strengths and limitations. *Pharmacoepidemiol Drug Saf* 2001; **10**: 341–344. doi:10.1002/pds.656.

5. Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med* 1965; **58**: 295–300.

6. Carstensen B, Witte DR, Friis S. Cancer occurrence in Danish diabetic patients: duration and insulin effects. *Diabetologia* 2012; **55**: 948–958. doi:10.1007/s00125-011-2381-4.

7. Colhoun HM, Livingstone SJ, Looker HC, *et al.* Hospitalised hip fracture risk with rosiglitazone and pioglitazone use compared with other glucose-lowering drugs. *Diabetologia* 2012; **55**: 2929–2937. doi:10.1007/s00125-012-2668-0.

8. Levin D, Bell S, Sund R, *et al.* Pioglitazone and bladder cancer risk: a multipopulation pooled, cumulative exposure analysis. *Diabetologia* 2015; **58**: 493–504. doi:10.1007/s00125-014-3456-9.

9. Pyörälä K, Pedersen TR, Kjekshus J, Faergeman O, Olsson AG, Thorgeirsson G. Cholesterol lowering with simvastatin improves prognosis of diabetic patients with coronary heart disease. A subgroup analysis of the Scandinavian Simvastatin Survival Study (4S). *Diabetes Care* 1997; **20**: 614–620.

10. Goldberg RB, Mellies MJ, Sacks FM, *et al.* Cardiovascular events and their reduction with pravastatin in diabetic and glucose-intolerant myocardial infarction survivors with average cholesterol levels: subgroup analyses in the cholesterol and recurrent events (CARE) trial. The Care Investigators. *Circulation* 1998; **98**: 2513–2519.

11. Colhoun HM, Betteridge DJ, Durrington PN, *et al.* Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. *Lancet Lond Engl* 2004; **364**: 685–696. doi:10.1016/S0140-6736(04)16895-5.

12. Cholesterol Treatment Trialists' (CTT) Collaborators, Kearney PM, Blackwell L, *et al.* Efficacy of cholesterol-lowering therapy in 18,686 people with diabetes in 14 randomised trials of statins: a meta-analysis. *Lancet Lond Engl* 2008; **371**: 117–125. doi:10.1016/S0140-6736(08)60104-X.

13. Hiatt WR. Observational Studies of Drug Safety — Aprotinin and the Absence of Transparency. *N Engl J Med* 2006; **355**: 2171–2173. doi:10.1056/NEJMp068252.

14. Szabo BG, Kadar B, Lenart KS, *et al.* Use of intravenous tigecycline in patients with severe Clostridium difficile infection: a retrospective observational cohort study. *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis* 2016. doi:10.1016/j.cmi.2016.08.017.

15. Bass SN, Bauer SR, Neuner EA, Lam SW. Comparison of treatment outcomes with vancomycin alone versus combination therapy in severe Clostridium difficile infection. *J Hosp Infect* 2013; **85**: 22–27. doi:10.1016/j.jhin.2012.12.019.

16. Byun J-I, Lee S-T, Jung K-H, *et al.* Effect of Immunotherapy on Seizure Outcome in Patients with Autoimmune Encephalitis: A Prospective Observational Registry Study. *PloS One* 2016; **11**: e0146455. doi:10.1371/journal.pone.0146455.

17. Muso E, Mune M, Hirano T, *et al.* A Prospective Observational Survey on the Long-Term Effect of LDL Apheresis on Drug-Resistant Nephrotic Syndrome. *Nephron Extra* 2015; **5**: 58–66. doi:10.1159/000437338.

18. Liang W, Zhao Y, Lee AH. A proxy outcome approach for causal effect in observational studies: a simulation study. *BioMed Res Int* 2014; **2014**: 872435. doi:10.1155/2014/872435.

19. Yu H, Yin L, Jiang X, *et al.* Effect of metformin on cancer risk and treatment outcome of prostate cancer: a meta-analysis of epidemiological observational studies. *PloS One* 2014; **9**: e116327. doi:10.1371/journal.pone.0116327.

20. Livingstone SJ, Looker HC, Hothersall EJ, *et al.* Risk of cardiovascular disease and total mortality in adults with type 1 diabetes: Scottish registry linkage study. *PLoS Med* 2012; **9**: e1001321. doi:10.1371/journal.pmed.1001321.

21. Maul A. A discrete time logistic regression model for analyzing censored survival data. *Environmetrics* 1994; **5**: 145–157. doi:10.1002/env.3170050205.

22. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet Lond Engl* 2002; **360**: 7–22. doi:10.1016/S0140-6736(02)09327-3.

23. Lewis JD, Ferrara A, Peng T, *et al.* Risk of bladder cancer among diabetic patients treated with pioglitazone: interim report of a longitudinal cohort study. *Diabetes Care* 2011; **34**: 916–922. doi:10.2337/dc10-1068.

24. Neumann A, Weill A, Ricordeau P, Fagot JP, Alla F, Allemand H. Pioglitazone and risk of bladder cancer among diabetic patients in France: a population-based cohort study. *Diabetologia* 2012; **55**: 1953–1962. doi:10.1007/s00125-012-2538-9.

25. Sanchez RG, Sanchez Gomez LM, Carmona L, Figuls MR, Cosp XB. Hormone replacement therapy for preventing cardiovascular disease in post-menopausal women. *Cochrane Database Syst Rev* 2005: CD002229. doi:10.1002/14651858.CD002229.pub2.

26. Miettinen OS. Components of the crude risk ratio. *Am J Epidemiol* 1972; **96**: 168–172.

27. Andersson C, Vaag A, Selmer C, *et al.* Risk of cancer in patients using glucose-lowering agents: a nationwide cohort study of 3.6 million people. *BMJ Open* 2012; **2**. doi:10.1136/bmjopen-2011-000433.

28. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol* 2008; **167**: 492–499. doi:10.1093/aje/kwm324.

29. Yang X, Chan JCN. Comment on: Suissa and Azoulay. Metformin and the risk of cancer: time-related biases in observational studies. Diabetes Care 2012;35:2665-2673. *Diabetes Care* 2013; **36**: e87. doi:10.2337/dc12-2561.

30. Suissa S, Azoulay L. Response to Yang and Chan. Metformin and the risk of cancer: time-related biases in observational studies. Diabetes care 2012;35:2665-2673. *Diabetes Care* 2013; **36**: e88. doi:10.2337/dc13-0133.

31. van Staa TP, Patel D, Gallagher AM, Bruin ML. Glucose-lowering agents and the patterns of risk for cancer: a study with the General Practice Research Database and secondary care data. *Diabetologia* 2012; **55**: 654–665. doi:10.1007/s00125-011-2390-3.

Table 1. Description of Study Population on July 1st 2007 (Study Mid-Point) by Ever-Never Statin Use.

	Statin user (N=90417) †	Non-statin (N=30254) †	p-value	Missing (%)
Age (years)	66.3 (57.4, 73.4)	65.4 (54.4, 75.4)	<0.001	0 (0)
Male (%)	45618 (50.5)	16442 (54.3)	<0.001	0 (0)
Ever smoker (%)	61694 (68.2)	19074 (63.0)	<0.001	0 (0)
Diabetes duration (years)	5.4 (2.7, 9.5)	4.4 (1.3, 8.6)	<0.001	0 (0)
Age at diagnosis (years)	58.6 (50.0, 67.0)	59 (48.2, 69.0)	<0.001	0 (0)
BMI (kg/m^2)	30.6 (27.1, 34.9)	30.1 (26.3, 34.9)	<0.001	8457 (7.1)
Systolic blood pressure (mmHg)	135.0 (126.0, 142.0)	136.0 (126.0, 144.0)	<0.001	2535 (2.1)
Diastolic blood pressure (mmHg)	77.0 (70.0, 81.0)	78.0 (70.0, 82.0)	<0.001	2535 (2.1)
Total cholesterol (mmol/L)	4.1 (3.6, 4.7)	4.5 (4.0, 5.1)	<0.001	4136 (3.4)
HbA_{1c} (mmol/L) ‡	54.1 (47.5, 65.0)	54.1 (46.4, 65.0)	0.075	3254 (2.7)
On anti-hypertensive drug (%)	68232 (75.5)	16381 (54.1)	<0.001	0 (0)
Estimated glomerular filtration rate ($ml/min/1.73m^2$)	66.8 (55.1, 79.2)	68.5 (56.5, 81.6)	<0.001	10728 (8.89)

† Median (IQR) unless otherwise indicated; ‡ HbA_{1c} – glycosylated haemoglobin A_{1c}

5

Table 2. Hazard Ratios for Incident CVD Using a Time-Updated Ever-Never Exposure Term Only.

	Model with ever-exposure term only				Model with terms for ever-exposure and cumulative exposure			
	Hazard Ratio	Lower 95% Confidence bound	Upper 95% Confidence bound	p-value	Hazard Ratio	Lower 95% Confidence bound	Upper 95% Confidence bound	p-value
Current age (years) †	1.04	1.04	1.05	<0.001	1.05	1.04	1.06	<0.001
Gender (female vs male)	0.79	0.77	0.80	<0.001	0.79	0.77	0.80	<0.001
Age at diabetes diagnosis (years)	0.98	0.97	0.98	<0.001	0.97	0.97	0.98	<0.001
Statins ever vs never exposure	1.13	1.10	1.16	<0.001	1.20	1.16	1.23	<0.001
Cumulative statin exposure (years)	-	-	-	-	0.97	0.97	0.98	<0.001

† Models included linear and quadratic terms for age

Table 3. Hazard ratios for incident CVD in models with cumulative statin exposure coded as discrete categories, illustrating effect of choice of baseline category. Models include sex, age at baseline, age at diabetes diagnosis, and time-updated calendar time.

	With baseline category as unexposed person-time intervals				With baseline category as exposed person-time intervals with cumulative exposure < 0.5 years			
Years of cumulative exposure	Hazard Ratio	Lower 95% Confidence bound	Upper 95% Confidence bound	p-value	Hazard Ratio	Lower 95% Confidence bound	Upper 95% Confidence bound	p-value
0 < exposure <=0.5	1.29	1.24	1.34	<0.000	NA	NA	NA	NA
0.5< exposure <=1.5	1.13	1.09	1.17	<0.000	0.88	0.84	0.92	<0.001
1.5< exposure <=2.5	1.07	1.03	1.12	0.001	0.83	0.79	0.87	<0.001
2.5< exposure <=3.5	1.09	1.04	1.14	<0.000	0.85	0.81	0.89	<0.001
3.5< exposure <=4.5	1.09	1.04	1.14	0.001	0.84	0.80	0.89	<0.001
4.5< exposure <= 5.5	1.06	1.01	1.12	0.023	0.83	0.78	0.88	<0.001
5.5< exposure <=6.5	1.04	0.98	1.11	0.186	0.81	0.76	0.86	<0.001
6.5< exposure <=7.5	1.05	0.98	1.13	0.157	0.82	0.76	0.88	<0.001
7.5< exposure <=8.5	1.07	0.95	1.20	0.245	0.83	0.74	0.94	0.002

5

Figure legend: Age-standardised rates of CVD by cumulative exposure to statins. The error bars indicate the 95% CI for the rates. The *x* axis shows cumulative years of exposure; the data point at $x = 0$ is for all unexposed person time-periods, and the other data points are for exposure categories $0 < x \leq 0.5$, $0.5 < x \leq 1$, $1 < x \leq 1.5$, $1.5 < x \leq 2$, $2 < x \leq 2.5$, $2.5 < x \leq 3$, $3 < x \leq 3.5$, $3.5 < x \leq 4$, $4 < x \leq 4.5$, $4.5 < x \leq 5$, $5 < x \leq 5.5$, $5.5 < x \leq 6$, and $6 < x \leq 6.5$. The rates in the initial exposed time period will reflect the starting rate and any effect of the drug within that time interval. To estimate the rate at the point of exposure a fitted line or curve through the points representing exposed person-time intervals would have to be extrapolated back to zero exposure i.e. the intercept on the Y-axis. The ratio of the observed rate in unexposed intervals and the fitted rate at the point of starting exposure *in those exposed* is the effect of confounding by indication or allocation bias. Where the time course effect is correctly modelled this is given by the hazard ratio for the ever never term in the model described.